BRIEF REPORT

More heads choose better than one: Group decision making can eliminate probability matching

Christin Schulze^{1,2} · Ben R. Newell²

Published online: 30 September 2015
© Psychonomic Society, Inc. 2015

Abstract Probability matching is a robust and common failure to adhere to normative predictions in sequential decision making. We show that this choice anomaly is nearly eradicated by gathering individual decision makers into small groups and asking the groups to decide. The group choice advantage emerged both when participants generated responses for an entire sequence of choices without outcome feedback (Exp. 1a) and when participants made trial-by-trial predictions with outcome feedback after each decision (Exp. 1b). We show that the dramatic improvement observed in group settings stands in stark contrast to a complete lack of effective solitary deliberation. These findings suggest a crucial role of group discussion in alleviating the impact of hasty intuitive responses in tasks better suited to careful deliberation.

Keywords Probability matching · Group decision making · Rational choice theory

Myriad everyday decisions are bound by limited resources and demand rapid and efficient processing. The heuristics-and-biases research tradition has established that fast, intuitive choices often yield effective outcomes in such situations, but sometimes fall prey to systematic errors under remarkably transparent circumstances (Gilovich, Griffin, & Kahneman,

2002). A paradigmatic example is *probability matching*—a striking error defined as selecting available choice options in proportion to the options’ relative success rates (for a review, see Vulkan, 2000).

Consider a standard repeated choice problem in which a fair, ten-sided die with seven red and three green sides is rolled many times (e.g., James & Koehler, 2011). Participants select a color for each roll and receive a fixed payment for each outcome they predict correctly. The likely anticipation that the die will come up in proportion to its color configuration has been argued to evoke probability-matching’s intuitive appeal (Koehler & James, 2009). That is, in light of this expectation, the corresponding strategy of choosing red for 70% and green for 30% of rolls may come to mind readily and hold the (elusive) promise of complete predictive success. Only by accepting the inevitability of the probabilistic process, however, can payoffs be maximized, by selecting the die’s dominant color exclusively—that is, by *probability maximizing*. Probability matching has been demonstrated with different paradigms, and similar proportions of participants match in probability-learning tasks (e.g., Shanks, Tunney, & McCarthy, 2002), in repeated-choice tasks with clearly stated outcome probabilities (e.g., Koehler & James, 2009; Newell, Koehler, James, Rakow, & van Ravenzwaaij, 2013), and when making a single response by selecting a global strategy for a sequence of choices (e.g., Newell & Rakow, 2007; West & Stanovich, 2003).

Why do people make this mistake? The “expectation matching” account outlined above offers an explanation commensurate with a broader theoretical framework suggesting that probability matching represents an intuitive impulse that can be corrected through appropriate deliberation (Koehler & James, 2009; Kogler & Kühberger, 2007; West & Stanovich, 2003). Supporting this account, it has been shown that

✉ Christin Schulze
cschulze@mpib-berlin.mpg.de

¹ Max Planck Institute for Human Development, Berlin, Germany

² University of New South Wales, Sydney, Australia

probability matching can be reduced by blocking the generation of sequence-wide outcome expectations (James & Koehler, 2011) and by portraying the task as a statistical test, thereby encouraging deliberation (Kogler & Kühberger, 2007). Moreover, maximizing is more prevalent in individuals with high cognitive (and arguably deliberative) reasoning abilities (West & Stanovich, 2003) and when applicable choice strategies are described prior to the task (Koehler & James, 2010; Newell et al., 2013).

Building on this research, we used a group decision making paradigm to augment deliberative capacity in repeated choice. The central goal of our study was to determine whether the enhanced cognitive resources of a small group of people, as compared to the cognitive capacity of each individual member, can reduce probability matching. Although groups evidently outperform the average individual on a variety of problems and decisions—for instance, rule induction problems (Laughlin, VanderStoep, & Hollingshead, 1991) and interactive strategic games (Kugler, Kausel, & Kocher, 2012)—groups' imperviousness to systematic cognitive biases is less robust. That is, groups both reduce and amplify individual biases. A key determinant of groups' capacity to overcome (or succumb to) individual bias is the nature of the group process (Kerr, MacCoun, & Kramer, 1996), which has been found to hinge on the demonstrability of the correct solution (Laughlin & Ellis, 1986). If the demonstrability is low, individual opinions need to be aggregated (e.g., via a majority decision process), which may exacerbate biases that are widespread among individuals. If the bias can be demonstrated, it may be overcome by the group on the basis of a single member's insight.

Here, we compared group and individual choice in a two-part study. In Experiment 1a, participants generated responses for an entire sequence of choices before receiving outcome feedback; in Experiment 1b, participants made trial-by-trial choices while receiving outcome feedback following each decision. This distinction between generating a sequence of decisions without feedback and making the same number of repeated choices with trial-by-trial feedback does not appear to affect people's (typically suboptimal) responses in individual choice (Newell & Rakow, 2007). We evaluated whether this finding would generalize to group choice or whether groups' performance hinges on the processes involved in planning decisions for an entire sequence of choices. That is, continuous outcome feedback might undermine the potential demonstrability of optimal maximization via the accumulation of "incorrect" predictions. Experiencing the inevitable errors implied by a maximizing strategy might lead to dissension in the group and a switch to a less optimal mode of responding. We discuss relevant group processes and their role for the demonstrability of maximizing as the correct solution, but focus primarily on evaluating groups' and individuals' choice strategies.

Experiment 1a

Group choice was compared to individual choice in two control conditions: individuals who started the task immediately and individuals who were afforded an amount of "deliberation time" equivalent to the time the groups had available to discuss prior to the choice task. This manipulation offered a direct control for the group choice context and allowed us to test whether the simple affordance of time to ponder strategies is sufficient to increase probability maximizing.

Method

Participants A total of 180 undergraduate students from the University of New South Wales (101 female, 79 male) with a mean age of 19.08 years ($SD = 2.56$) participated in Experiment 1a in exchange for course credit. Additionally, participants earned a performance-based payment, and earnings ranged from \$2.10 to \$4.10 (1 AUD $\approx$ 0.90 USD at the time the experiment was conducted).

Design and procedure Participants were invited to the experiment in groups of four people. Fifteen groups of four participants each were randomly assigned to one of three conditions: individual control, individual with time for deliberation, or group choice. Participants in both individual choice conditions were seated at computers in four separate rooms, and instructions were presented on the computer screen; participants in the group choice condition were seated at the same computer in a larger room, and instructions were administered on paper and read aloud by the experimenter. Following the presentation of instructions, participants in the group choice and individual time-for-deliberation conditions were given 10 min to discuss/deliberate strategies for completing the choice task, whereas participants in the individual control condition started the task immediately.

The choice task involved a fair, virtual, ten-sided die with seven red and three green sides (counterbalanced across groups of four people for red/green majority outcomes) that was rolled 50 times on a computer screen. The instructions clearly indicated that the die was not biased or loaded in any way, but was fair in the sense that each side had an equal chance of turning up once the die was rolled. Participants were asked to generate a sequence of 50 bets that would maximize their earnings by selecting a color for each roll on the screen. Figure 1 illustrates the choice part of the task interface in Experiment 1a (panel a) and contrasts it with the trial-by-trial response mode of Experiment 1b (panel b). The figure shows that all bets were placed before any rolls of the die occurred; that is, participants had to decide on a color for each roll before seeing any outcomes. Once the sequence of bets had been generated, the computer rolled the die automatically, and participants observed the outcome of each roll on the screen.

a Experiment 1a. Participants generated an entire choice sequence before receiving feedback.

Round	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Your Bet	RED	RED	RED	RED	RED	RED	RED	RED	RED	RED	RED	RED	RED	RED	RED	RED	RED	RED	RED	RED	RED	RED	RED	RED	RED

Round	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
Your Bet	RED	RED	RED	RED	RED	RED	RED	RED	RED

b Experiment 1b. Participants made trial-by-trial choices with feedback after each click.

Fig. 1 Illustrative comparison of choice task interfaces between experiments. In both experiments, participants were asked to make predictions about the rolls of a die described to them as having seven red and three green sides (colors counterbalanced). The upper panel shows how participants generated responses for an entire sequence of

choices by selecting colors on the computer screen, before receiving outcome feedback, in Experiment 1a; the lower panel illustrates how participants made trial-by-trial choices by repeatedly selecting buttons on the screen, followed by outcome feedback after each click, in Experiment 1b

Correct bets were rewarded with ten cents for each person—irrespective of whether participants completed the task as part of a group or individually—and each person's ensuing payoff was updated with each roll on the screen. Participants were encouraged to earn as much money as possible. The primary dependent measure was participants' (groups' or individuals') proportion of choices to the more probable—that is, dominant color—outcome.

Following the choice task, a questionnaire was presented that included queries about participants' strategy use during the choice task (adapted from Koehler & James, 2010), the Cognitive Reflection Test (CRT; Frederick, 2005), the Berlin Numeracy Test (BNT; Cokely, Galesic, Schulz, Ghazal, & Garcia-Retamero, 2012), and measures of self-reported mathematical ability and education.¹

Results

Following the methodological precedent from research on group choice in intellectual problem solving tasks, we compared the performance of the four-person groups to that of the best, second-best, third-best, and fourth-best of four independent participants in each individual choice condition

(Laughlin, Bonner, & Miner, 2002; Laughlin et al., 1991). Across all rolls of the die, participants' dominant color choice proportions were defined as the highest, second-highest, third-highest, and fourth-highest choice proportion within each set of four independent individuals. Where two or more individuals returned identical choice proportions, we assigned rank positions randomly (see, e.g., Laughlin et al., 1991). In addition to conventional methods of hypothesis testing, we conducted default Bayesian analyses of variance (ANOVAs; Rouder, Morey, Speckman, & Province, 2012), default Bayesian *t* tests (Rouder, Speckman, Sun, Morey, & Iverson, 2009), and default Bayesian hypothesis tests for correlations (Wetzels & Wagenmakers, 2012). We report Bayes factors (*BF*) that quantify the strength of the evidence in favor of the presence of an effect.²

Figure 2 compares the groups' dominant color choice proportions to those of the best, second-best, third-best, and fourth-best performing individuals in each individual choice

¹ We found only limited evidence that group choice affected the strategy evaluations after the choice task. In Experiment 1a, participants who had made their choices as part of a group more often correctly identified maximizing as the strategy with the highest expected payoff (90%) than did participants who had made their choices individually with or without time for deliberation (70% and 73%, respectively). This trend, however, was not clear cut for all questionnaire items and did not replicate in Experiment 1b. Therefore, we do not consider these analyses further.

² All Bayesian analyses were carried out in R. For the Bayesian ANOVAs and *t* tests, we used the *anovaBF* and *ttestBF* functions included in the *BayesFactor* package (v.0.9.9; Morey & Rouder, 2014) with their respective default settings, with one exception: For the *anovaBF* function, the number of Monte Carlo samples used to estimate *BF*s was increased to 50,000. For the Bayesian ANOVAs, we computed *BF*s by contrasting the performance of a model including the relevant effect to one omitting that effect. Follow-up comparisons were made by carrying out Bayesian *t* tests. For Bayesian *t* tests and Bayesian correlations, *BF*s in favor of the alternative hypothesis are reported. A *BF* of 5, for instance, suggests that the data are five times more likely to have occurred under the model assuming the relevant effect than under a model omitting this effect, whereas a *BF* of 0.20 indicates that the data are five times more likely to have occurred under the model omitting the relevant effect than under a model including this effect.

Fig. 2 Proportions of dominant color choices for groups and for the best, second-best, third-best, and fourth-best performing individuals in each individual choice condition in Experiment 1a. Bar graphs plot mean $\pm$ standard error choice proportions across groups/individuals. Small

squares/circles and diamonds plot the choice proportion of one group/individual in the respective decision making condition. Probability matching is indicated by the dashed line at .70

condition. For the highest-performing individuals, choice proportions did not differ between decision making conditions; in fact, the Bayesian analysis provided evidence in favor of the absence of an effect, $F(2, 42) = 0.29$, $p = .749$, $\eta_p^2 = .014$, $BF = 0.20$. However, we found significant differences between groups and individuals at the second, $F(2, 42) = 4.74$, $p = .014$, $\eta_p^2 = .184$, $BF = 4.03$; third, $F(2, 42) = 28.78$, $p < .001$, $\eta_p^2 = .578$, $BF = 7.73 \times 10^5$; and fourth, $F(2, 42) = 70.00$, $p < .001$, $\eta_p^2 = .769$, $BF = 1.21 \times 10^{11}$, choice ranks. Scheffé-based conventional post-hoc analyses and follow-up Bayesian t tests showed that groups selected the dominant color significantly more often than the second-best ($p = .022$, $BF = 8.87$), third-best ($p < .001$, $BF = 4.99 \times 10^4$), and fourth-best ($p < .001$, $BF = 6.75 \times 10^9$) individuals who started the task immediately. Similarly, groups performed (marginally) better than the second-best ($p = .073$, $BF = 5.06$), third-best ($p < .001$, $BF = 2.46 \times 10^5$), and fourth-best ($p < .001$, $BF = 3.31 \times 10^7$) individuals who were given time to deliberate strategies. The participants in both individual decision making conditions returned comparable choice proportions at all choice ranks (all p s $\geq .634$ and all BF s ≤ 0.45). The similar levels of choice performance in both individual decision making conditions may have stemmed from participants in the individual control condition compensating for their reduced level of cognitive processing time by spending more time deliberating strategies during the choice task. To examine this possibility, we compared task completion times between the decision making conditions but found no significant differences, $F(2, 132) = 1.37$, $p = .258$, $\eta_p^2 = .020$, $BF = 0.24$. On average, participants in the individual control, time-for-deliberation, and group choice condition spent 113.43 s ($SD = 45.08$), 109.30 s ($SD = 41.11$), and 93.27 s ($SD = 33.64$), respectively, on completing the task.

The full distribution of groups' and individuals' choice proportions is displayed alongside aggregate choice in Fig. 2. We classified these individual-level responses as

probability matching (allocating choices within 5% of the average outcome probability—i.e., $70\% \pm 5\%$) or probability maximizing (selecting the dominant color on at least 95% of rolls; see, e.g., Schulze, van Ravenzwaaij, & Newell, 2015). No single group of decision makers probability matched. Instead, the majority of the four-person groups (73%) adhered to optimal probability maximizing. By contrast, individual decision makers showed equal propensities toward probability matching and maximizing when we analyzed the data from all choice ranks; 35% of the participants in both individual choice conditions adopted either strategy.

People's performance on the CRT has been shown to correlate positively with accuracy in binary choice (e.g., Koehler & James, 2010). We confirmed these reports and found positive correlations between CRT scores and participants' dominant color choice proportions in the individual condition with time for deliberation, $r(58) = .610$, $p < .001$, $BF = 6.00 \times 10^4$, and in the individual control condition, $r(58) = .310$, $p = .016$, $BF = 1.82$. Controlling for numeracy, as indicated by BNT performance, self-reported mathematical ability, and degree of mathematical education, this relationship remained intact for the individual condition with time for deliberation, $r(55) = .286$, $p = .031$, $BF = 1.89$, but not for the individual control condition, $r(55) = .215$, $p = .108$, $BF = 0.81$. By contrast, no such relationships existed for groups. We examined the correlation between groups' dominant color choice proportions and individual group member CRT scores as well as group CRT characteristics, such as the highest CRT score of the best group member and the sum of individual CRT scores within a group. We found no significant correlations between groups' dominant color choice proportions and CRT measures, irrespective of whether or not we controlled for numeracy, self-reported mathematical ability, and mathematical education for each group member or the sum/best member score within a group, respectively (all p s $\geq .238$ and all BF s ≤ 0.36). It is important to note, however, that groups' dominant color choice proportions

were close to ceiling and showed little variability (see Fig. 2), which may have hampered the detectability of links between choice performance and CRT scores.

Experiment 1b

We compared group and individual choices in another experiment in which participants made trial-by-trial choices and received outcome feedback *following each decision*. That is, rather than generating a sequence of 50 bets before receiving any feedback, the participants in Experiment 1b made one bet, rolled the die, saw the outcome, and then made another bet for each of the 50 rolls. Because we had observed no effect of giving individuals time to deliberate strategies prior to the task in Experiment 1a, we reduced the experimental design to two between-subjects conditions (group choice with 10 min for discussion vs. individual choice without predetermined time for deliberation) in Experiment 1b.

Method

Participants A total of 120 participants (63 female, 57 male) were recruited via the subject pool of the Max Planck Institute for Human Development. The mean age was 24.63 years ($SD = 3.26$). As in Experiment 1a, participants earned a performance-based payment (increased to €0.20 per correct choice; earnings ranged from €4.20 to €8.40, 1 EUR $\approx$ 1.12 USD at the time the experiment was conducted), and all participants received an additional flat fee of €5.

Choice task and procedure The choice task and procedure replicated those of Experiment 1a with the following exceptions. Individuals and groups made 50 choices by clicking repeatedly on one of two buttons, labeled “RED” and “GREEN,” on the computer screen (see Fig. 1b). Following each choice, the computer rolled the die and participants observed the outcome of the roll on the screen. That is, choices were made on the basis of a description of the problem as well as of trial-by-trial outcome experience. Choice trials were self-paced, and participants made a response via mouse click to advance to the next trial after observing the outcome feedback. Following the choice task, each participant in the group choice condition answered an additional questionnaire about (1) which group members had contributed to the group solution (by identifying none, one, two, three, or all group members via preallocated letters); (2) the process through which the group reached a decision (based on the decision schemes typically identified in group choice research, as summarized in Table 1; see Kerr et al., 1996); and (3) whether the group’s strategy was the same or different from the strategy that the participant would have used if the game had been played alone.

Table 1 Self-reported group choice processes for all 15 interacting four-person groups, sorted by overall group choice solution (i.e., the proportion of dominant color choices)

Choice	Unanimity	Truth Wins	Equiprobability	Majority
1	++++			
1	++++			
1	++++			
1	++++			
1	+++Ø			
1	+++	Ø		
1	+	++Ø		
1		++ØØ		
.98		+ØØØ		
.98	+++Ø			
.94	++	+		Ø
.94	+	+	+Ø	
.84	+++Ø			
.80		+++	Ø	
.74		++ØØ		
Count	35	21	3	1

Symbols indicate the selected process of each group member and whether that participant would have used the same (+) or a different (Ø) strategy, if the game had been played alone. Participants self-identified group processes by selecting the statement that best described their group’s decision process from among the following options: (a) The group members wanted to follow the same strategy, there was no need for discussion, and a consensus was found immediately (*unanimity*); or, the group members wanted to follow different strategies and (b) the majority opinion was followed after voting (*majority*), (c) the best strategy that was proposed by a group member was followed (truth wins), or (d) any proposed strategy was followed (*equiprobability*).

Results

Figure 3 compares groups’ dominant color choice proportions to those of the best, second-best, third-best, and fourth-best performing individuals in the individual choice condition. We found significant differences in choice proportions between group choice and individual choice at different rank levels, $F(4, 70) = 13.70$, $p < .001$, $\eta_p^2 = .439$, $BF = 4.75 \times 10^5$. Comparing group choice proportions to individual choices at each rank level, Dunnett *t*-test-based conventional post-hoc analyses and follow-up Bayesian *t*tests showed that groups performed as well as the best ($p = .614$, $BF = 0.80$) and second-best ($p = .999$, $BF = 0.35$) individuals, but selected the dominant color significantly more often than the third-best ($p = .019$, $BF = 3.48$) and fourth-best ($p < .001$, $BF = 341.48$) individuals. Classifying individual groups’ and participants’ choices as probability matching or maximizing (see Exp. 1a) revealed that only one group of decision makers probability matched; the majority of the four-person groups (67%) probability maximized correctly. As compared to Experiment 1a, however, more participants in the individual

Fig. 3 From left to right, proportions of dominant color choices for groups and for the best, second-best, third-best, and fourth-best performing individuals in the individual control condition in Experiment 1b. Bar graphs plot mean $\pm$ standard error choice proportions across groups/individuals. Small squares/circles plot the choice proportion of one group/individual in the respective decision making condition. Probability matching is indicated by the dashed line at .70

choice condition probability maximized (43%), and fewer probability matched (10%).

Again, we compared the task completion times between decision making conditions but found no significant differences, $t(15.44) = 1.64$, $p = .122$, $BF = 3.68$, although the Bayesian analysis suggested that individual participants spent slightly less time on the task ($M = 257.03$ s, $SD = 30.63$) than did the groups ($M = 286.60$ s, $SD = 68.24$). Thus, individuals did not appear to compensate for their lack of explicit deliberation time during the task.

Turning to the questionnaire data, we again examined the relationship between participants' CRT scores and their dominant color choice proportions during the choice task. Because in Experiment 1b we used a different subject pool with potentially higher previous exposure to common experimental measures in judgment and decision making research, we asked participants to indicate how many items of the CRT and the BNT they had previously encountered (none, some, or all). Participants' prior exposure to the test items was positively correlated with their performance on the CRT, $r(118) = .404$, $p < .001$, $BF = 2.45 \times 10^3$, but not on the BNT, $r(118) = .049$, $p = .593$, $BF = 0.08$. Therefore, all following correlational analyses were computed while controlling for self-indicated prior knowledge of CRT test items for each person or the sum/highest knowledge within a group, respectively. For participants in the individual choice condition, we, again, found a positive correlation between CRT scores and dominant color choice proportions, $r(57) = .356$, $p = .006$, $BF = 7.18$, which did not hold when controlling for BNT score, self-reported mathematical ability, and degree of mathematical education, $r(54) = .219$, $p = .105$, $BF = 0.88$. For participants in the group choice condition, we, again, found no relationship between groups' dominant color choice proportions and the individual-level CRT scores, the highest CRT score of the best

group member, or the sum of the individual CRT scores within a group, irrespective of whether or not we controlled for numeracy, self-reported mathematical ability, and mathematical education for each group member or the sum/best-member score within a group (all $ps \geq .127$ and all $BFs \leq 1.28$).

Examining the self-reported group processes, we found large intragroup agreement on all three questionnaire items. In the majority of groups, three or more members agreed upon the process through which a solution was reached (87%). For the majority of participants in the group choice condition, three or more group members agreed upon whether or not that person had contributed to the group solution (75%), and most group members indicated that they would have followed the same strategy as their group, if the task had been solitary (75%). Table 1 summarizes participants' responses and shows that most group members (58%) indicated that their group followed a unanimous group process, in which almost all members would have used the same strategy in a solitary game. Approximately one third of participants indicated that the group reached consensus by a "truth wins" procedure, and many of these participants indicated that they would not have come up with the same strategy if they had played alone. By contrast, only a single person indicated that the group had reached consensus by following the majority opinion.

Discussion

Individuals are often notoriously bad at adhering to principles of rational choice when making repeated decisions (Vulkan, 2000). We demonstrated that groups perform as well as their best individual members and better than almost everyone else in sequential binary choice. In fact, assembling people into groups nearly eradicated inferior probability matching—a success rate seldom achieved in individual binary choice, at which at least some people persistently probability match over hundreds of handsomely remunerated choice trials, and despite sometimes heavy-handed motivation schemes (Shanks et al., 2002). Groups performed remarkably well when participants generated responses for an entire sequence of choices without outcome feedback (Exp. 1a), as well as when they made trial-by-trial predictions with outcome feedback after each decision (Exp. 1b). Yet, in the latter situation, group choice was slightly less optimal: One group probability matched, and the number of probability maximizing groups decreased by one. As compared to this small loss in group performance, we found larger differences in individual choice between experiments, which likely were associated with differences in the respective subject pools. Whereas the participants in Experiment 1a had participated in return for credit in an introductory undergraduate psychology course, the subject pool for Experiment 1b was more diverse, potentially more highly incentivized, and indicated considerable prior exposure

to common experimental measures in judgment and decision making research.

What drives groups' superior performance in sequential choice? The self-reported group processes elicited from participants after the choice task in Experiment 1b suggest that probability matching indeed represents a demonstrably inferior strategy. That is, most group members indicated that their group either followed a "truth wins" decision process and implemented the most effective strategy proposed by one of the members—even if they would not have thought of the same solution by themselves—or was unanimous in identifying the correct solution. Only a single person perceived the process as a majority vote. Thus, a central contribution of this article is the finding that the erroneous nature of probability matching can be demonstrated and overridden through small group discussions—even if a large portion of the group's members initially held different beliefs. Nonetheless, it is possible that not all groups were able to demonstrate the correct solution to their members and that the trial-by-trial outcome feedback provided in Experiment 1b slightly undermined the demonstrability of optimal maximization, thus leading to the small loss in group performance that we observed.

Whether probability matching's biased nature is, in principle, demonstrable was far from obvious a priori. In fact, sequential choice paradigms have previously been characterized by the *absence* of a demonstrably correct solution, because only outcome probabilities, not correct trial-by-trial outcomes, can be demonstrated before responses are made (Laughlin & Ellis, 1986). Moreover, research on related sequential choice tasks also solvable by selecting the strategy with the highest expected value has produced mixed results. Charness, Kami, and Levin (2007), for instance, found that groups violate the principle of stochastic dominance in risky choice significantly less often than individuals. Davis, Hornik, and Hornseth (1970), on the other hand, rejected only the highest-expected-value process model as an adequate account of group decisions throughout a sequential choice task and observed virtually no differences between groups' and individuals' choice strategies. Contrasting sequential decisions from experience in static and dynamic choice settings, Lejarraga, Lejarraga, and Gonzalez (2014) showed that groups outperform individuals in static settings but lose their advantage when the environment changes unexpectedly. Therefore, an interesting avenue for future research would be to explore the role of small-group discussion for effective choice in probability-learning or dynamic choice settings. In the former, a group advantage may still hold despite the absence of initial knowledge about the relative risks associated with each choice option because, if the group appreciates that outcome probabilities may be unequal, the group members can discuss the correct solution irrespective of the actual probabilities. The group advantage could then be realized analogously, once

the members have learned which option yields payment with higher odds.

The near-optimal effect of group deliberation that we observed in the present set of experiments is particularly compelling in light of the absolute lack of effective solitary deliberation. The extent to which this group achievement requires at least one member to recognize matching's inferiority a priori, however, remains an open question—we did not examine strategy evaluations *prior* to the task, to avoid confounding effects of strategy availability (see Koehler & James, 2010). Nevertheless, measuring cognitive individual differences revealed that group choice afforded an advantage over individual choice beyond the simple sum of each group member's cognitive and mathematical abilities. Thus, group discussion may have facilitated collective processing beyond individual member capabilities—analogous to the "assembly bonus effects" observed in intellectual problem solving (e.g., Laughlin et al., 2002).

Probability matching exemplifies situations in which initial intuitions need to be overcome by careful deliberation. We conclude that such situations can be improved through group discussion, and point toward the potential relevance of these findings for surmounting intuitive fallacies in myriad real-life contexts—for instance, in juries, boards of directors, committees, and families.

Author note This research was supported by Australian Research Council grants to B.R.N. (Nos. DP110100797 and FT110100151).

References

- Charness, G., Kami, E., & Levin, D. (2007). Individual and group decision making under risk: An experimental study of Bayesian updating and violations of first-order stochastic dominance. *Journal of Risk and Uncertainty*, 35, 129–148.
- Cokely, E. T., Galesic, M., Schulz, E., Ghazal, S., & Garcia-Retamero, R. (2012). Measuring risk literacy: The Berlin Numeracy Test. *Judgment and Decision Making*, 7, 25–47.
- Davis, J. H., Hornik, J. A., & Hornseth, J. P. (1970). Group decision schemes and strategy preferences in a sequential response task. *Journal of Personality and Social Psychology*, 15, 397–408.
- Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19, 25–42.
- Gilovich, T., Griffin, D., & Kahneman, D. (Eds.). (2002). *Heuristics and biases: The psychology of intuitive judgment*. Cambridge, UK: Cambridge University Press.
- James, G., & Koehler, D. J. (2011). Banking on a bad bet: Probability matching in risky choice is linked to expectation generation. *Psychological Science*, 22, 707–711. doi:10.1177/0956797611407933
- Kerr, N. L., MacCoun, R. J., & Kramer, G. P. (1996). Bias in judgment: Comparing individuals and groups. *Psychological Review*, 103, 687–719.
- Koehler, D. J., & James, G. (2009). Probability matching in choice under uncertainty: Intuition versus deliberation. *Cognition*, 113, 123–127.
- Koehler, D. J., & James, G. (2010). Probability matching and strategy availability. *Memory & Cognition*, 38, 667–676. doi:10.3758/MC.38.6.667

- Kogler, C., & Kühberger, A. (2007). Dual process theories: A key for understanding the diversification bias? *Journal of Risk and Uncertainty*, 34, 145–154.
- Kugler, T., Kausel, E. E., & Kocher, M. G. (2012). Are groups more rational than individuals? A review of interactive decision making in groups. *Wiley Interdisciplinary Reviews: Cognitive Science*, 3, 471–482. doi:10.1002/wcs.1184
- Laughlin, P. R., Bonner, B. L., & Miner, A. G. (2002). Groups perform better than the best individuals on Letters-to-Numbers problems. *Organizational Behavior and Human Decision Processes*, 88, 605–620.
- Laughlin, P. R., & Ellis, A. L. (1986). Demonstrability and social combination processes on mathematical intellectual tasks. *Journal of Experimental Social Psychology*, 22, 177–189.
- Laughlin, P. R., VanderStoep, S. W., & Hollingshead, A. B. (1991). Collective versus individual induction: Recognition of truth, rejection of error, and collective information processing. *Journal of Personality and Social Psychology*, 61, 50–67.
- Lejarraga, T., Lejarraga, J., & Gonzalez, C. (2014). Decisions from experience: How groups and individuals adapt to change. *Memory & Cognition*, 42, 1384–1397. doi:10.3758/s13421-014-0445-7
- Morey, R. D., & Rouder, J. N. (2014). BayesFactor (Version 0.9.9). Retrieved from <http://bayesfactorpcl.r-forge.r-project.org>
- Newell, B. R., Koehler, D. J., James, G., Rakow, T., & van Ravenzwaaij, D. (2013). Probability matching in risky choice: The interplay of feedback and strategy availability. *Memory & Cognition*, 41, 329–338. doi:10.3758/s13421-012-0268-3
- Newell, B. R., & Rakow, T. (2007). The role of experience in decisions from description. *Psychonomic Bulletin & Review*, 14, 1133–1139. doi:10.3758/BF03193102
- Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, 56, 356–374. doi:10.1016/j.jmp.2012.08.001
- Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian *t* tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, 16, 225–237. doi:10.3758/PBR.16.2.225
- Schulze, C., van Ravenzwaaij, D., & Newell, B. R. (2015). Of matchers and maximizers: How competition shapes choice under risk and uncertainty. *Cognitive Psychology*, 78, 78–98.
- Shanks, D. R., Tunney, R. J., & McCarthy, J. D. (2002). A re-examination of probability matching and rational choice. *Journal of Behavioral Decision Making*, 15, 233–250.
- Vulkan, N. (2000). An economist's perspective on probability matching. *Journal of Economic Surveys*, 14, 101–118.
- West, R. F., & Stanovich, K. E. (2003). Is probability matching smart? Associations between probabilistic choices and cognitive ability. *Memory & Cognition*, 31, 243–251. doi:10.3758/BF03194383
- Wetzels, R., & Wagenmakers, E.-J. (2012). A default Bayesian hypothesis test for correlations and partial correlations. *Psychonomic Bulletin & Review*, 19, 1057–1064. doi:10.3758/s13423-012-0295-x